There are 4 different brains, with some changes:

Default: Has the next rewards 🡪

* The target arrives to the goal = +5.0
* Every frame that the target isn’t in the goal = (-1f / agentParameters.maxStep)

BigReward: Has the same rewards as Default, except in 🡪

* The target arrives to the goal = +10.0 (A bigger reward for completing the task)

BigTime: Has the same rewards as Default, except in 🡪

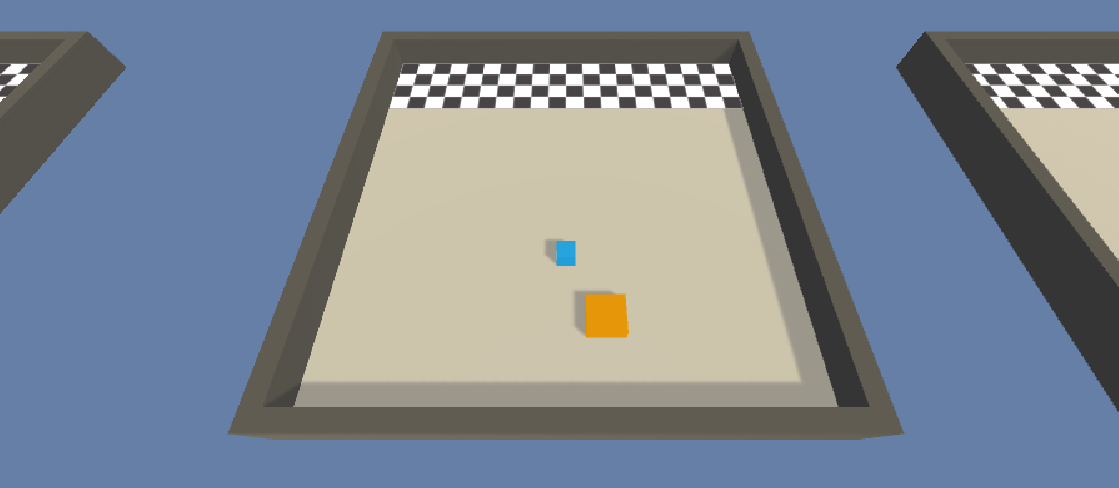
* Every frame that the target isn’t in the goal = (-2f / agentParameters.maxStep) (The most slower is, less reward will have)

OurLearning: Has the rewards of Default and two additional rewards 🡪

* If the agent is at a max distance of 3.0f with the target, every frame = + 0.001f
* If the target is at a max distance of 1.0f with a wall, every frame = -0.01f

We tested those brains into 3 different maps:

Map 1:



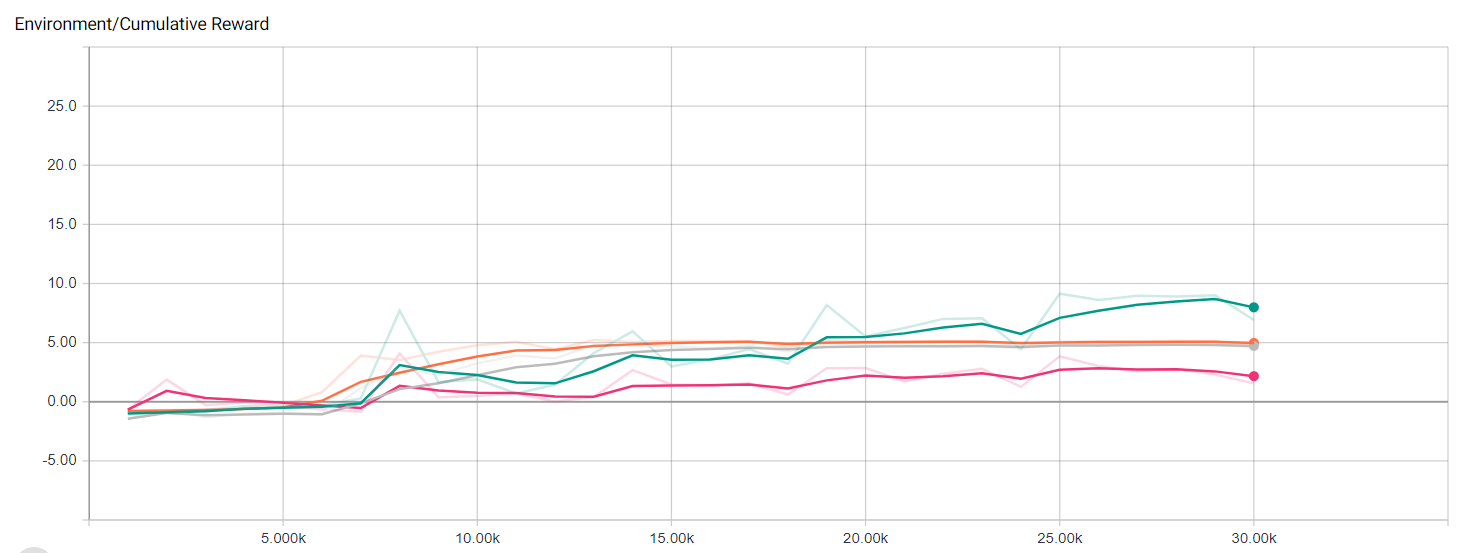
Default 🡪 Pink

BigReward 🡪 Green/Blue

BigTime 🡪 Grey

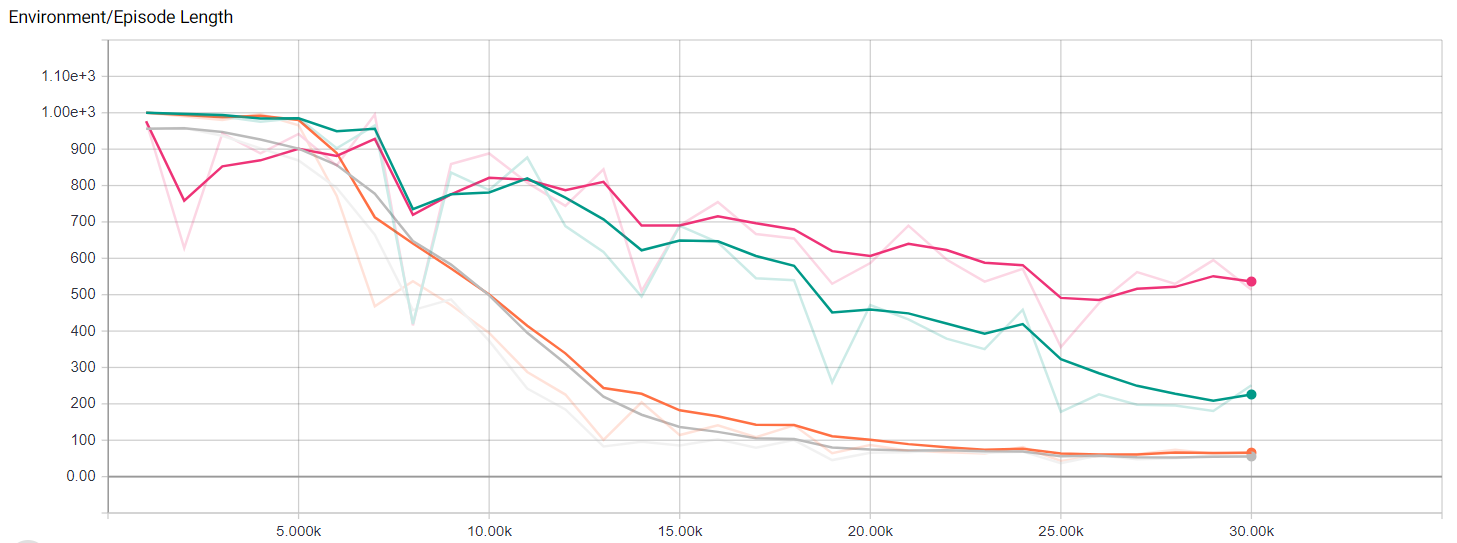
OurLearning 🡪Orange

Cumulative reward:



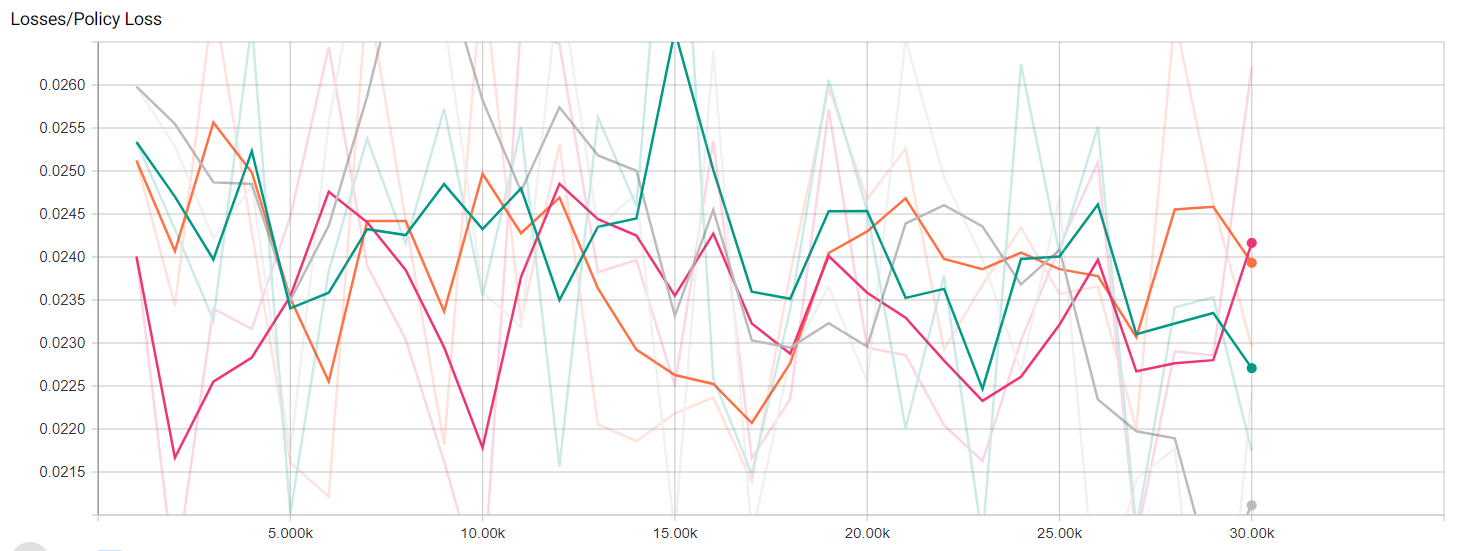
The better is the BigReward and the worst the default

Episode length:



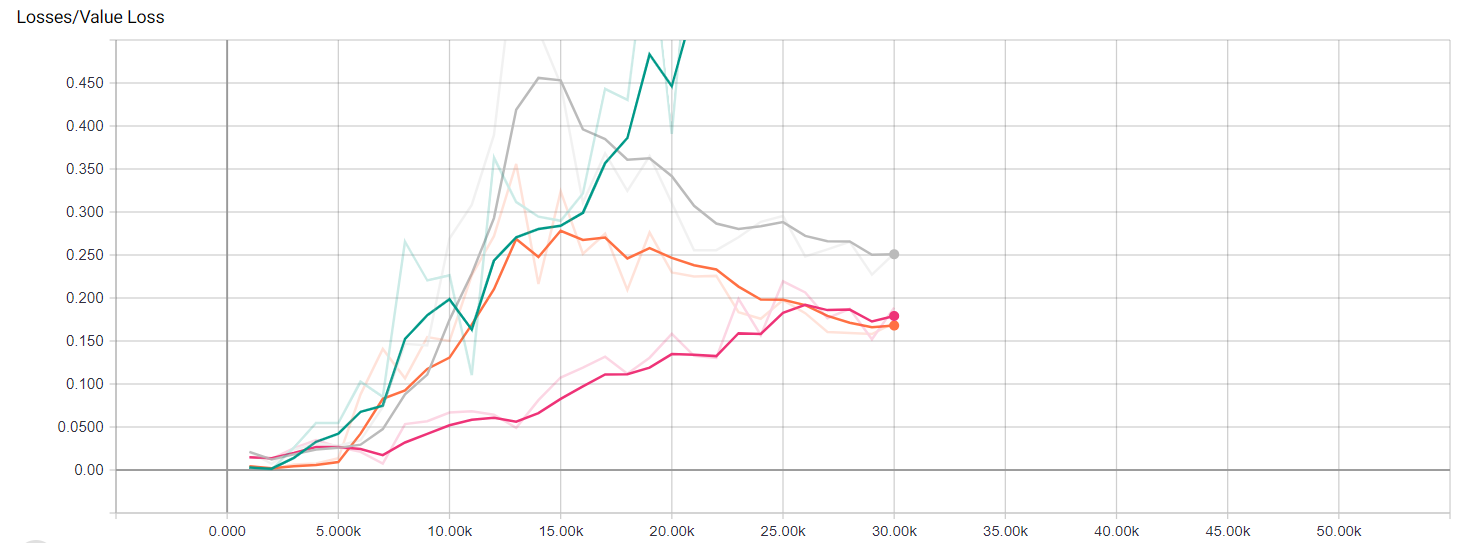
The better is OurLearning and the worst the default

Policy loss:



It’s very chaotic, and is hard to choose the better one…

Value loss:



The worst by far is the BigReward and the best is OurLearning

Map 2:



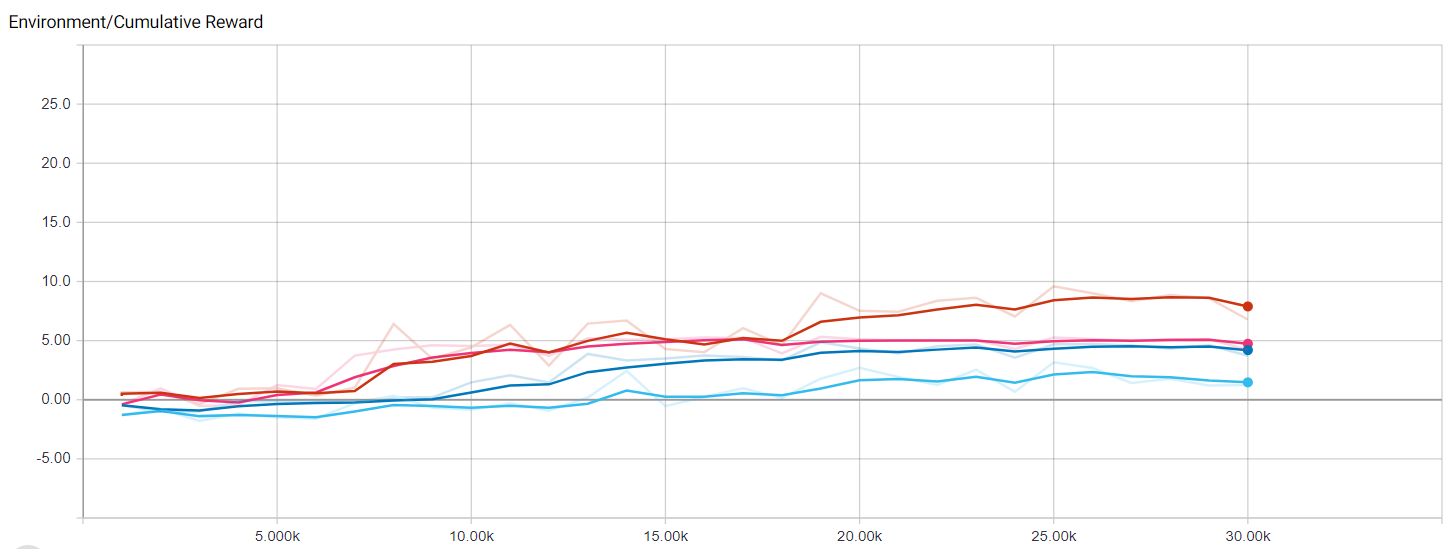
Default 🡪 Blue

BigReward 🡪 Red

BigTime 🡪 Green/Blue

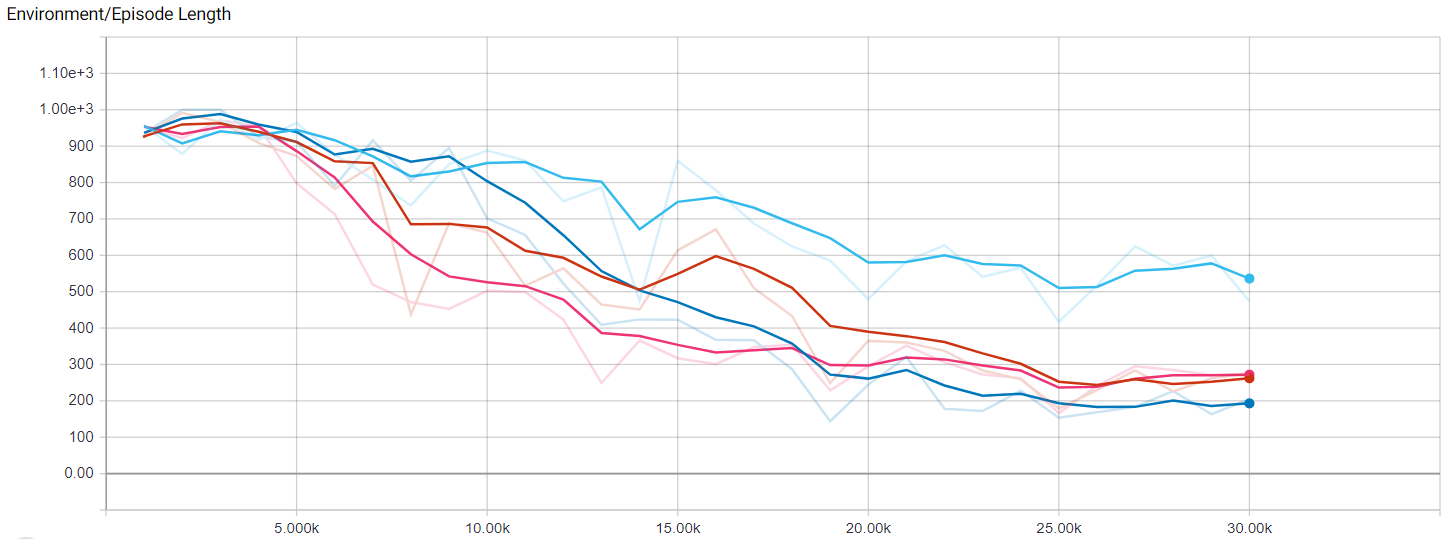
OurLearning 🡪Pink

Cumulative loss:



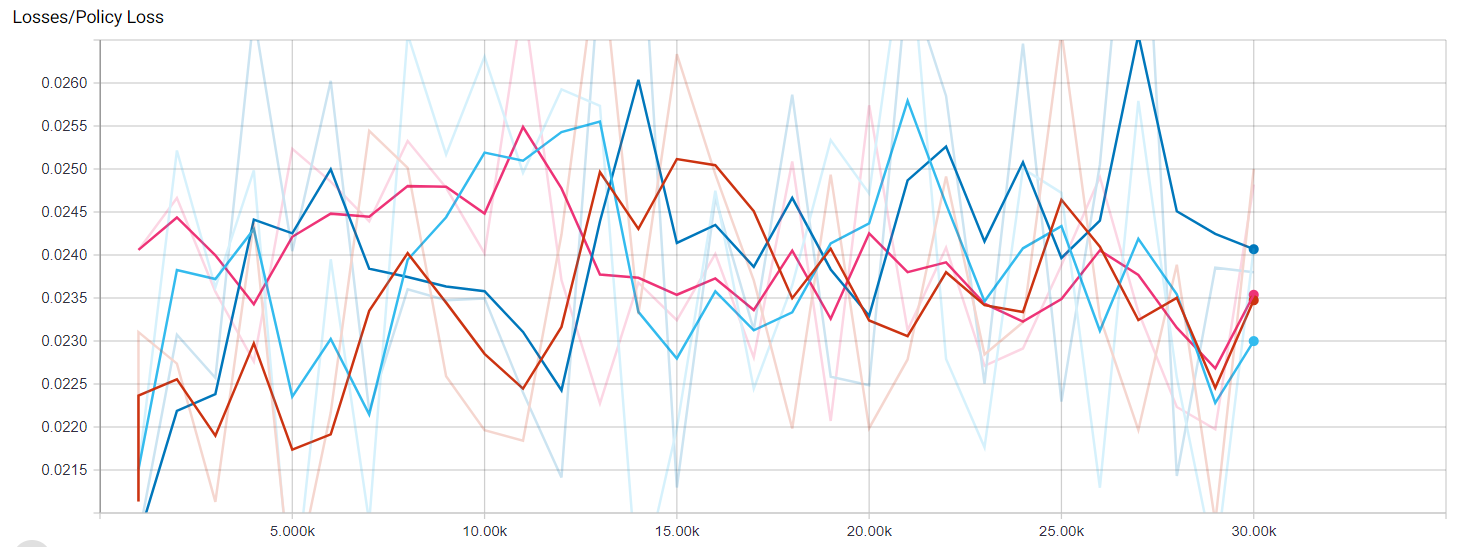
The best is the BigReward, and the worst the BigTime

Episode length:



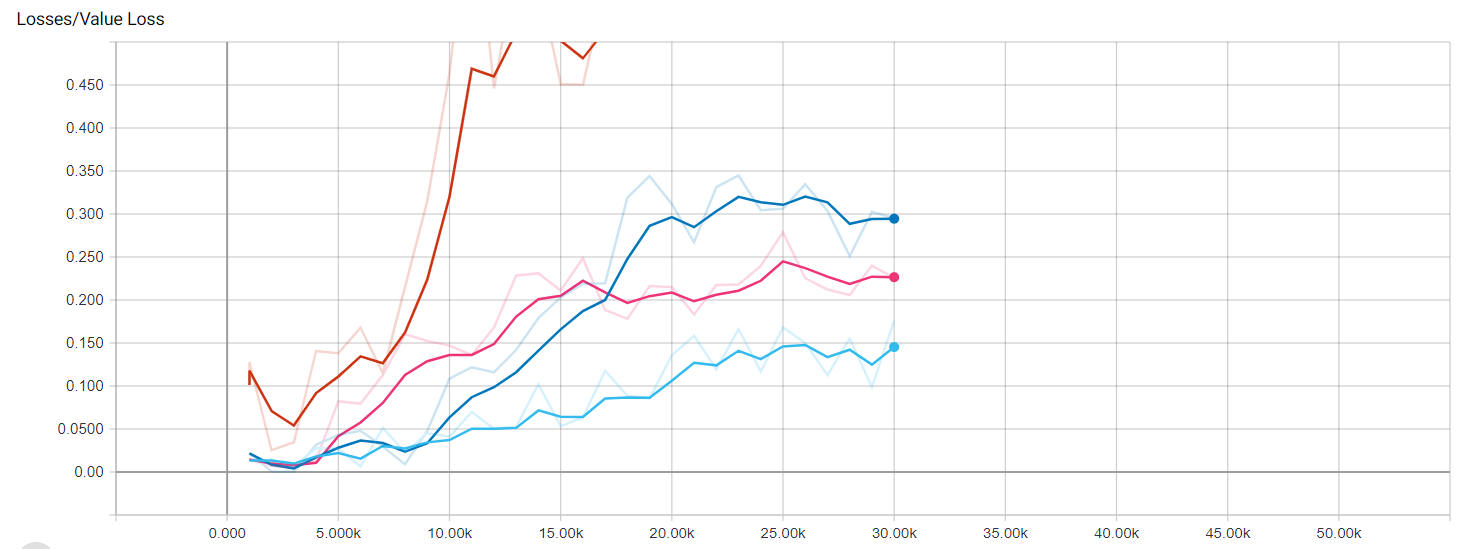
The best is the default, and the worst the BigTime

Policy Loss:



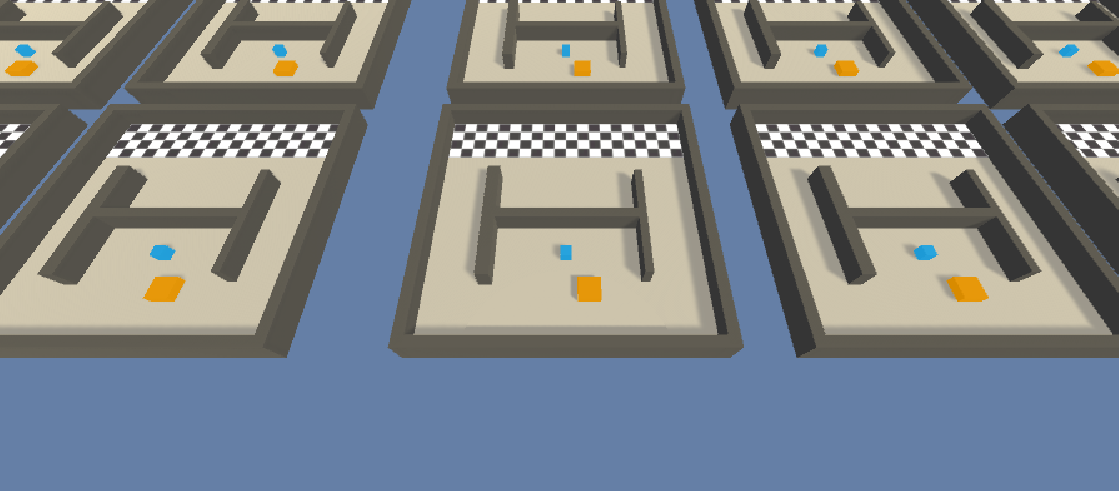
Its very chaotic and it’s hard to choose the best and the worst

Value loss:



The best one is the BigTime and the worst by far is the BigReward

Map 3:



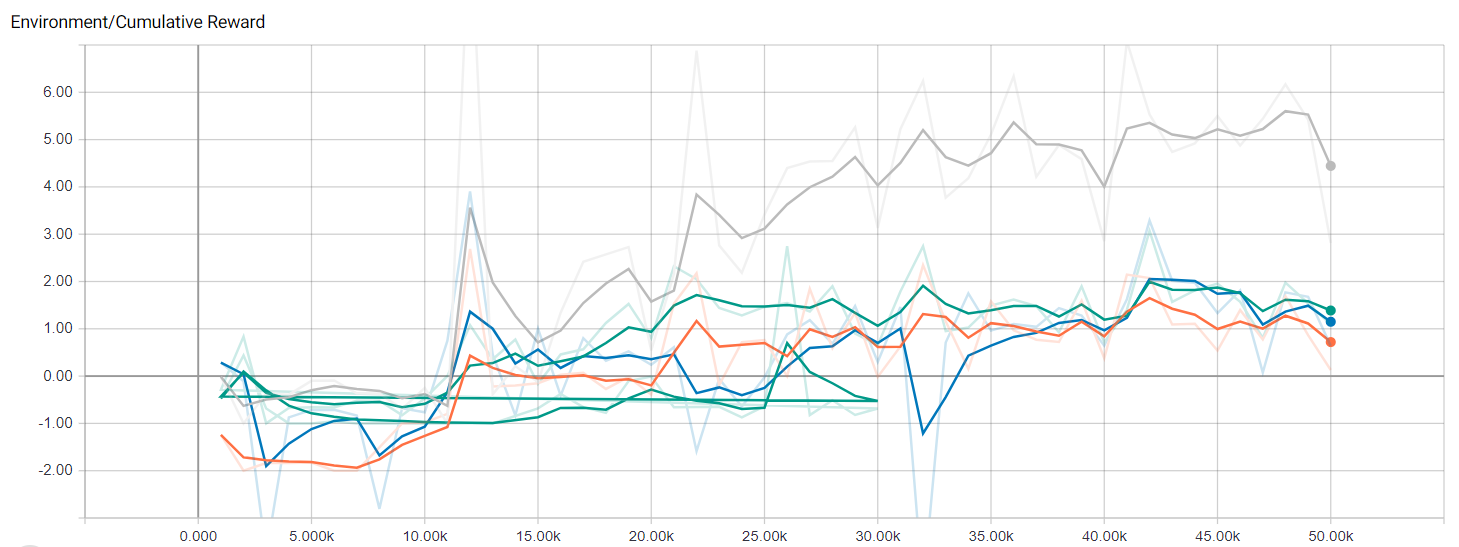
Default 🡪 Blue/Green

BigReward 🡪 Grey

BigTime 🡪 Orange

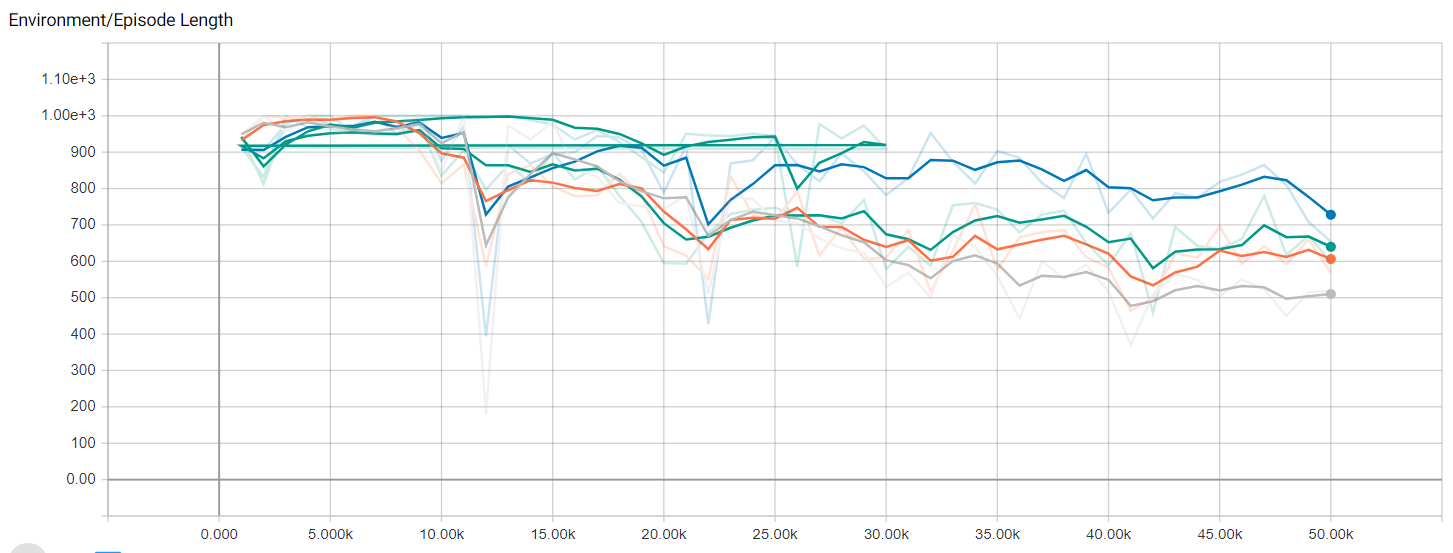
OurLearning 🡪Blue

Cumulative loss:



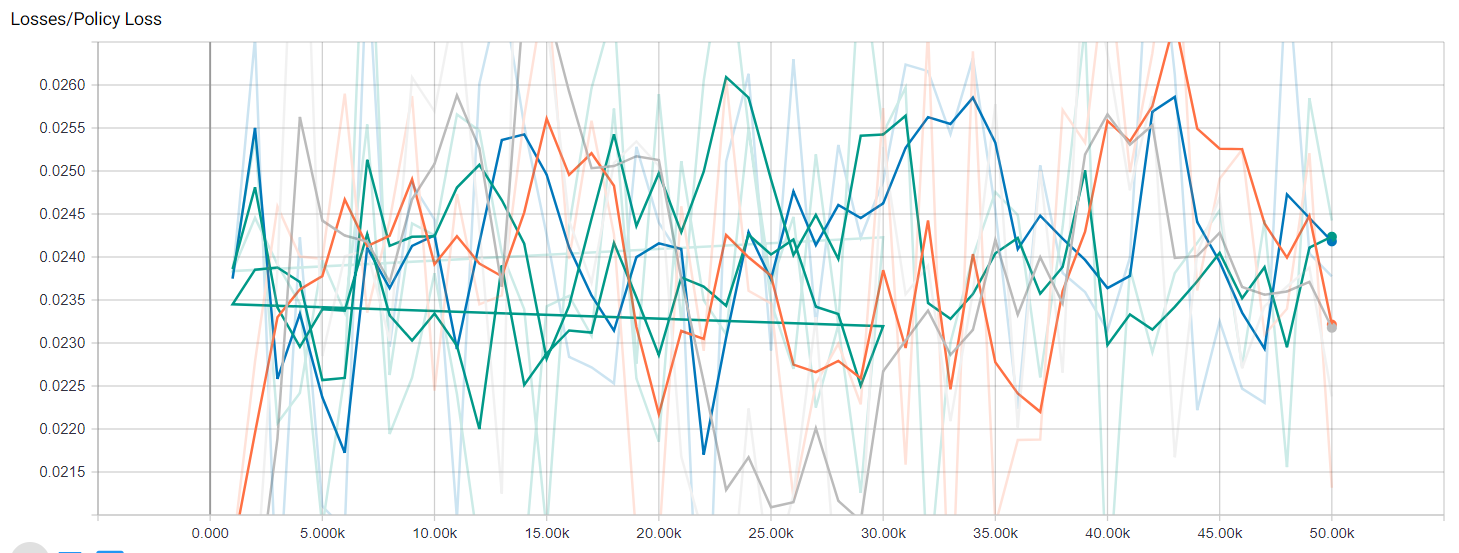
The best one is the BigReward and the worst the BigTime

Episode length:



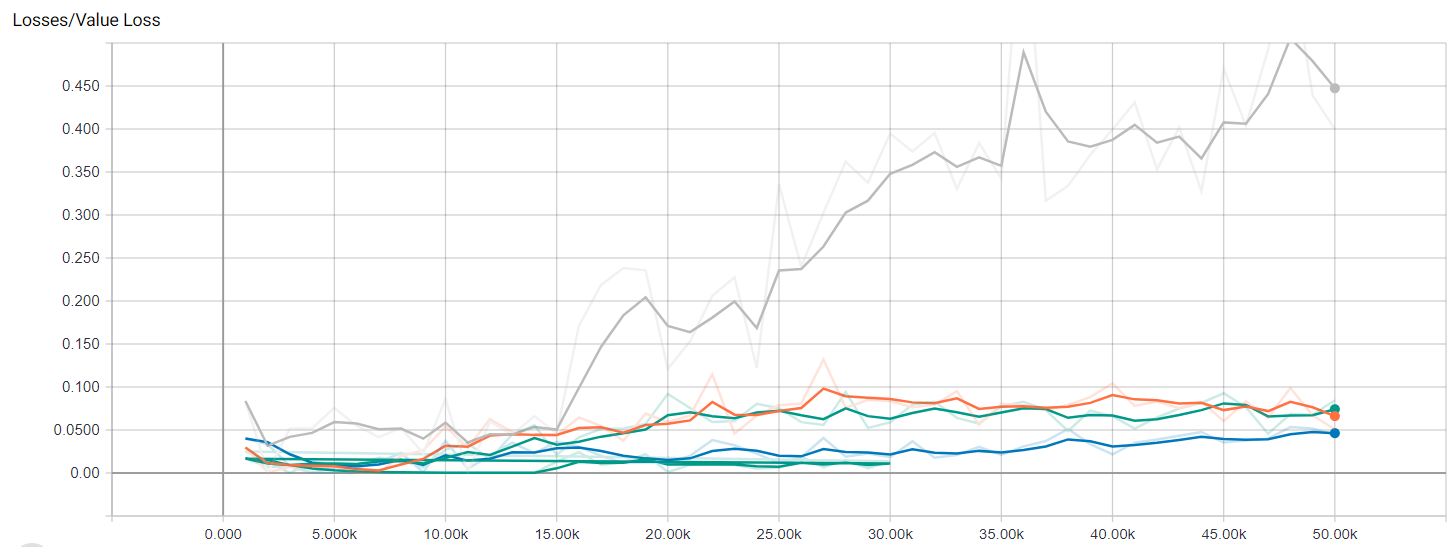
The best is the BigReward and the worst OurLearning

Policy loss:



Its very chaotic and it’s hard to choose the best and the worst

Value loss:



The best one is OurLearning and the worst by far, BigReward

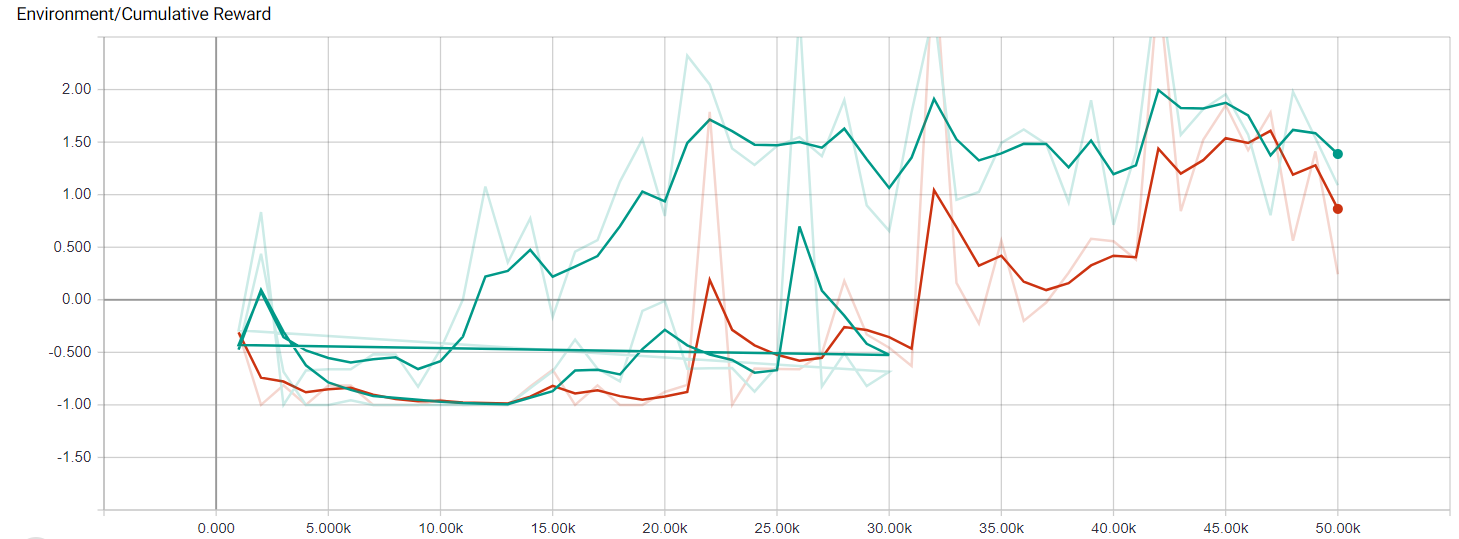
When we put each brain already trained in the maps, the one that seems to work best is OurLearning

**Imitation Learning:**

We trained a IL brain in 50.000 steps in Map3.

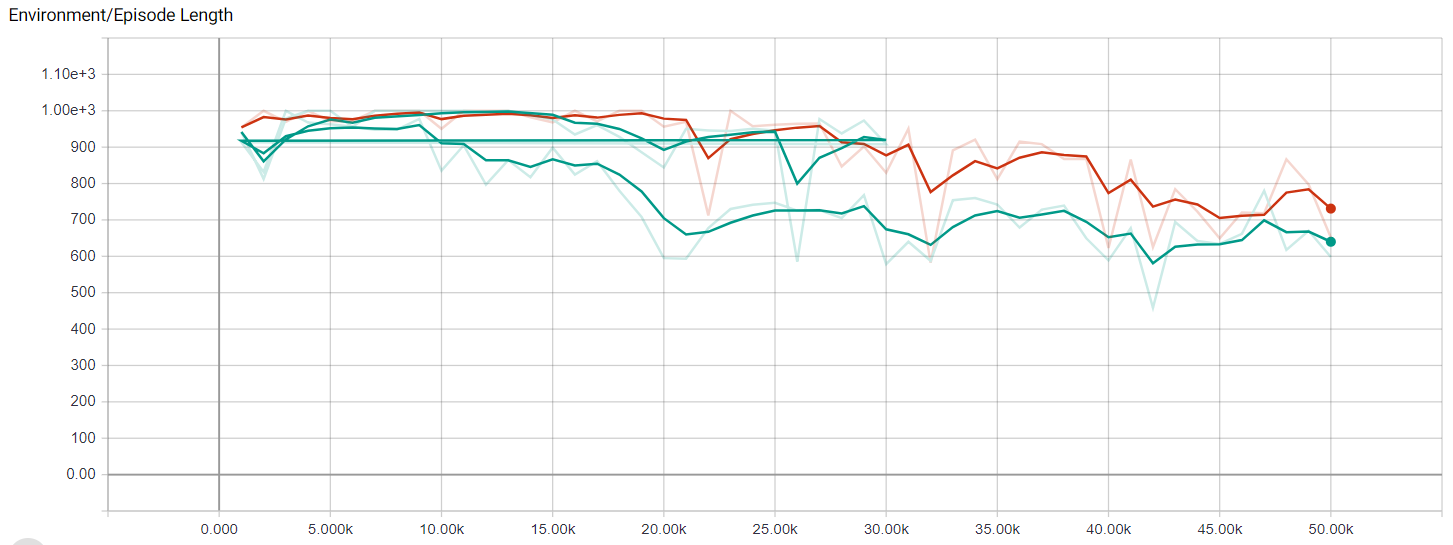
The IL train is the red and the Default is the blue

Cumulative Reward



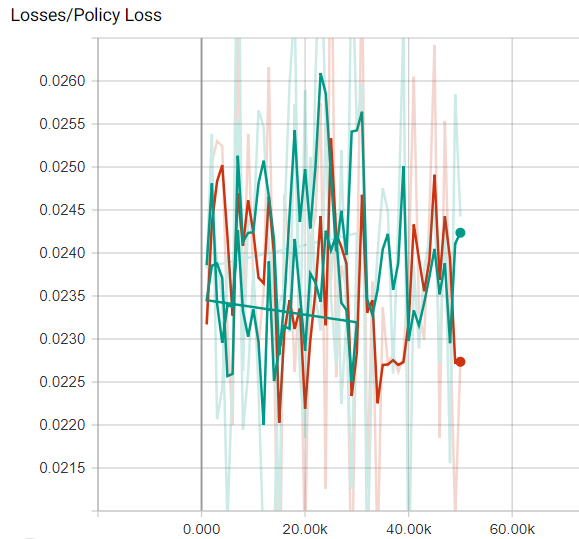
The Default is a bit better

Episode Lenght



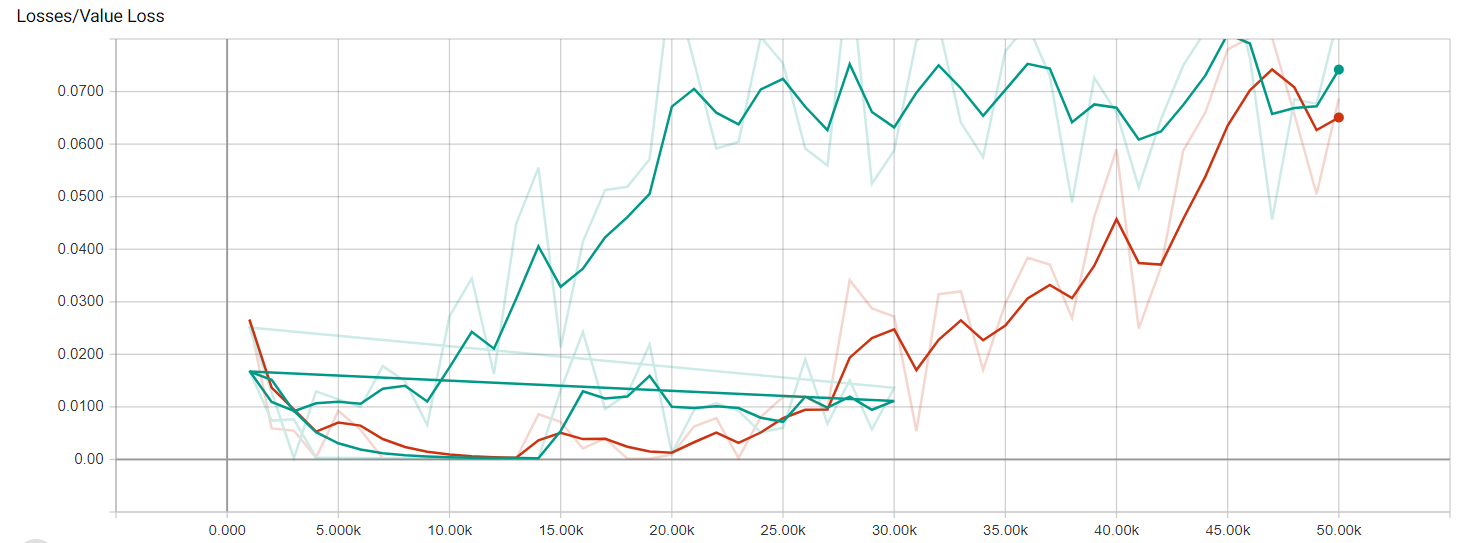
The Default is a bit better than the IL

Policy Loss



Its very chaotic, so its hard tho know the better one

Value loss



The IL Train is better than the Default

Conclusion: The IL is not as good as the Default, but this is positive because it needs to fail to seem more human